

Zero Attracting PNLMS Algorithm and Its Convergence in Mean

Rajib Lochan Das, *Member, IEEE*, and Mrityunjoy Chakraborty, *Senior Member, IEEE*

Abstract

The proportionate normalized least mean square (PNLMS) algorithm and its variants are by far the most popular adaptive filters that are used to identify sparse systems. The convergence speed of the PNLMS algorithm, though very high initially, however, slows down at a later stage, even becoming worse than sparsity agnostic adaptive filters like the NLMS. In this paper, we address this problem by introducing a carefully constructed l_1 norm (of the coefficients) penalty in the PNLMS cost function which favors sparsity. This results in certain zero attractor terms in the PNLMS weight update equation which help in the shrinkage of the coefficients, especially the inactive taps, thereby arresting the slowing down of convergence and also producing lesser steady state excess mean square error (EMSE). We also carry out the convergence analysis (in mean) of the proposed algorithm.

Index Terms

Sparse Adaptive Filter, PNLMS Algorithm, RZA-NLMS algorithm, convergence speed, steady state performance.

I. INTRODUCTION

IN real life, there exist many examples of systems that have a sparse impulse response, having a few significant non-zero elements (called active taps) amidst several zero or insignificant elements (called inactive taps). One example of such systems is the network echo canceller [1]- [2], which uses both packet-switched and circuit-switched components and has a total echo response of about 64-128 ms duration out of which the “active” region spans a duration of only 8-12 ms, while the remaining “inactive” part accounts for bulk delay due to network loading, encoding and jitter buffer delays. Another example is the acoustic echo generated due to coupling between microphone and loudspeaker in hands free mobile telephony, where the sparsity of the acoustic channel impulse response varies with the loudspeaker-microphone distance [3]. Other well known examples of sparse systems include HDTV where clusters of dominant echoes arrive after long periods of silence [4], wireless multipath channels which, on most of the occasions, have only a few clusters of significant paths [5], and underwater acoustic channels where the various multipath components caused by reflections off the sea surface and sea bed have long intermediate delays [6]. The last decade witnessed a flurry of research activities [7] that sought to develop sparsity aware adaptive filters which can exploit the a priori knowledge of the sparseness of the system and thus enjoy improved identification performance. The first and foremost in this category is the proportionate normalized LMS (PNLMS) algorithm [8] which achieves faster initial convergence by deploying different step sizes for different weights, with each one made proportional to the magnitude of the corresponding weight estimate. The convergence rate of the PNLMS algorithm, however, slows down at a later stage of the iteration and becomes even worse than a sparsity agnostic algorithm like the NLMS [9]. This problem was later addressed in several of its variants like the improved PNLMS (i.e. IPNLMS) algorithm [11], composite proportionate and normalized LMS (i.e. CPNLMS) algorithm [10] and mu law PNLMS (i.e. MPNLMS) algorithm [13]. These algorithms improve the transient response (i.e. convergence speed) of the PNLMS algorithm for identifying sparse systems. However, all of them yield almost same steady-state excess mean square error (EMSE) performance as produced by the PNLMS. The need to improve both transient and steady-state performance subsequently led to several variable step-size (VSS), proportionate type algorithms [14]- [16].

In this paper, drawing ideas from [17]- [18], we aim to improve the performance of the PNLMS algorithm further, by introducing a carefully constructed l_1 norm (of the coefficients) penalty in the PNLMS cost function which favors sparsity¹. This results in a modified PNLMS update equation with a zero attractor for all the taps, named as the Zero-Attracting PNLMS (ZA-PNLMS) algorithm. The zero attractors help in the shrinkage of the coefficients which is particularly desirable for the inactive taps, thereby giving rise to lesser steady state EMSE for sparse systems. Further, by drawing the inactive taps towards zero, the zero attractors help in arresting the sluggishness of the convergence of the PNLMS algorithm that takes place at a later stage of the iteration, caused by the diminishing effective step sizes of the inactive taps. We show this by presenting a detailed convergence analysis of the proposed algorithm, which is, however, a very daunting task, especially due to the presence of a so-called gain matrix and also the zero attractors in the update equation. To overcome the challenges posed by them, we deploy a transform domain equivalent model of the proposed algorithm and separately, an elegant scheme of angular discretization of continuous valued random vectors proposed earlier in [22].

R. L. Das is with the Department of Electronics and Communication Engineering, Dr. B. C. Roy Engineering College, Durgapur, INDIA. M. Chakraborty is with the Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur, INDIA. (e.mail : rajib.das.iit@gmail.com; mrityun@ece.iitkgp.ernet.in).

¹Some preliminary results of this work were earlier presented by the authors at ISCAS 2014 [20].

II. PROPOSED ALGORITHM

Consider a PNLMS based adaptive filter that takes $x(n)$ as input and updates a L tap coefficient vector $\mathbf{w}(n) = [w_0(n), w_1(n), \dots, w_{L-1}(n)]^T$ as [8],

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu \mathbf{G}(n) \mathbf{x}(n) e(n)}{\mathbf{x}^T(n) \mathbf{G}(n) \mathbf{x}(n) + \delta_P}, \quad (1)$$

where $\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-L+1)]^T$ is the input regressor vector, $\mathbf{G}(n)$ is a diagonal matrix that modifies the step size of each tap, μ is the overall step size, δ_P is a regularization parameter and $e(n) = d(n) - \mathbf{w}^T(n) \mathbf{x}(n)$ is the filter output error, with $d(n)$ denoting the so-called desired response. In the system identification problem under consideration, $d(n)$ is the observed system output, given as $d(n) = \mathbf{w}_{opt}^T \mathbf{x}(n) + v(n)$, where \mathbf{w}_{opt} is the system impulse response vector (supposed to be sparse), $x(n)$ is the system input and $v(n)$ is an observation noise which is assumed to be white with variance σ_v^2 and independent of $x(m)$ for all n and m .

The matrix $\mathbf{G}(n)$ is evaluated as

$$\mathbf{G}(n) = \text{diag}(g_0(n), g_1(n), \dots, g_{L-1}(n)), \quad (2)$$

where,

$$g_l(n) = \frac{\gamma_l(n)}{\sum_{i=0}^{L-1} \gamma_i(n)}, \quad 0 \leq l \leq L-1, \quad (3)$$

with

$$\gamma_l(n) = \max[\rho_g \max[\delta, |w_0(n)|, \dots, |w_{L-1}(n)|], |w_l(n)|]. \quad (4)$$

The parameter δ is an initialization parameter that helps to prevent stalling of the weight updating at the initial stage when all the taps are initialized to zero. Similarly, if an individual tap weight becomes very small, to avoid stalling of the corresponding weight update recursion, the respective $\gamma_l(n)$ is taken as a small fraction (given by the constant ρ_g) of the largest tap magnitude. By providing separate effective step size $\mu g_l(n)$ to each l -th tap where $g_l(n)$ is broadly proportional to $|w_l(n)|$, the PNLMS algorithm achieves higher rate of convergence initially, caused primarily by the active taps. At a later stage, however, the convergence slows down considerably, being controlled primarily by the numerically dominant inactive taps that have progressively diminishing effective step sizes [11], [13].

It has recently been shown [21] that the PNLMS weight update recursion (i.e., Eq. (1)) can be obtained by minimizing the cost function $\|\mathbf{w}(n+1) - \mathbf{w}(n)\|_{\mathbf{G}^{-1}(n)}^2$ subject to the condition $d(n) - \mathbf{w}^T(n+1) \mathbf{x}(n) = 0$ (the notation $\|\mathbf{x}\|_{\mathbf{A}}^2$ indicates the generalized inner product $\mathbf{x}^T \mathbf{A} \mathbf{x}$). In order to derive the ZA-PNLMS algorithm, following [17], we add an l_1 norm penalty $\gamma \|\mathbf{G}^{-1}(n) \mathbf{w}(n+1)\|_1$ to the above cost function, where γ is a very very small constant. Note that unlike [17], we have, however, used a generalized form of l_1 norm penalty here which scales the elements of $\mathbf{w}(n+1)$ by $\mathbf{G}^{-1}(n)$ first before taking the l_1 norm (the above scaling makes the l_1 norm penalty governed primarily by the inactive taps). The above constrained optimization problem may then be stated as:

$$\min_{\mathbf{w}(n+1)} \|\mathbf{w}(n+1) - \mathbf{w}(n)\|_{\mathbf{G}^{-1}(n)}^2 + \gamma \|\mathbf{G}^{-1}(n) \mathbf{w}(n+1)\|_1 \quad (5)$$

subject to $d(n) - \mathbf{w}^T(n+1) \mathbf{x}(n) = 0$, where the short form notation “ \mathbf{G}^{-1} ” is used to indicate $\mathbf{G}^{-1}(n)$. Using Lagrange multiplier λ , this amounts to minimizing the cost function $J(n+1) = \|\mathbf{w}(n+1) - \mathbf{w}(n)\|_{\mathbf{G}^{-1}(n)}^2 + \gamma \|\mathbf{G}^{-1}(n) \mathbf{w}(n+1)\|_1 + \lambda(d(n) - \mathbf{w}^T(n+1) \mathbf{x}(n))$. Setting $\partial J(n+1)/\partial \mathbf{w}(n+1) = 0$, one obtains,

$$\mathbf{w}(n+1) = \mathbf{w}(n) - [\gamma \text{sgn}(\mathbf{w}(n+1)) - \lambda \mathbf{G}(n) \mathbf{x}(n)] \quad (6)$$

where $\text{sgn}(\cdot)$ is the well known signum function, i.e., $\text{sgn}(x) = 1$ ($x > 0$), 0 ($x = 0$), -1 ($x < 0$). Premultiplying both the LHS and the RHS of (6) by $\mathbf{x}^T(n)$ and using the condition $d(n) - \mathbf{w}^T(n+1) \mathbf{x}(n) = 0$, one obtains,

$$\lambda = \frac{e(n) + \gamma \mathbf{x}^T(n) \text{sgn}(\mathbf{w}(n+1))}{\mathbf{x}^T(n) \mathbf{G}(n) \mathbf{x}(n)}. \quad (7)$$

Substituting (7) in (6), we have,

$$\begin{aligned} \mathbf{w}(n+1) &= \mathbf{w}(n) + \frac{e(n) \mathbf{G}(n) \mathbf{x}(n)}{\mathbf{x}^T(n) \mathbf{G}(n) \mathbf{x}(n)} \\ &\quad - \gamma \left[\mathbf{I} - \frac{\mathbf{x}(n) \mathbf{x}^T(n) \mathbf{G}(n)}{\mathbf{x}^T(n) \mathbf{G}(n) \mathbf{x}(n)} \right] \text{sgn}(\mathbf{w}(n+1)). \end{aligned} \quad (8)$$

Note that the above equation does not provide the desired weight update relation, as the R.H.S. contains the unknown term $\text{sgn}(\mathbf{w}(n+1))$. In order to obtain a feasible weight update equation, we approximate $\text{sgn}(\mathbf{w}(n+1))$ by an estimate, namely, $\text{sgn}(\mathbf{w}(n))$ which is known. This is based on the assumption that most of the weights do not undergo change of sign as they

get updated. This assumption may not, however, appear to be a very accurate one, especially for the inactive taps that fluctuate around zero value in the steady state. Nevertheless, an analysis of the proposed algorithm, as given later in this paper, shows that this approximation does not have any serious effect on the convergence behavior of the proposed algorithm. Apart from this, we also observe that in (8), elements of the matrix $\frac{\mathbf{x}(n)\mathbf{x}^T(n)\mathbf{G}(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)}$ have magnitudes much less than 1, especially for large order filters, and thus, this term can be neglected in comparison to \mathbf{I} .

From above and introducing the algorithm step size μ and a regularization parameter δ_P in (8), for a large order adaptive filter, one then obtains the following weight update equation :

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu e(n)\mathbf{G}(n)\mathbf{x}(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n) + \delta_P} - \rho \text{sgn}(\mathbf{w}(n)) \quad (9)$$

where $\rho = \mu\gamma$.

Eq. (9) provides the weight update relation for the proposed ZA-PNLMS algorithm, where the second term on the R.H.S. is the usual PNLMS update term while the last term, i.e., $\rho \text{sgn}(\mathbf{w}(n))$ is the so-called zero attractor. The zero attractor adds $-\rho \text{sgn}(w_j(n))$ to $w_j(n)$ and thus helps in its shrinkage to zero. Ideally, the zero attraction should be confined only to the inactive taps, which means that the proposed ZA-PNLMS algorithm will perform particularly well for systems that are highly sparse, but its performance may degrade as the number of active taps increases. In such cases, Eq. (9) may be further refined by applying the reweighting concept [17] to it. For this, we replace the l_1 regularization term $\|\mathbf{G}^{-1}\mathbf{w}(n+1)\|_1$ in (5) by a log-sum penalty $\sum_{i=1}^L \frac{1}{g_i(n)} \log(1 + |w_i(n+1)|/\epsilon)$ where $g_i(n)$ is the i -th diagonal element of $\mathbf{G}(n)$ and ϵ is a small constant. Following the same steps as used above to derive the ZA-PNLMS algorithm, one can then obtain the RZA-PNLMS weight update equation as given by

$$\begin{aligned} w_i(n+1) &= w_i(n) + \frac{\mu g_i(n)x(n-i+1)e(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n) + \delta_P} \\ &\quad - \rho \frac{\text{sgn}(w_i(n))}{1 + \varepsilon |w_i(n)|}, \quad i = 0, 1, \dots, L-1, \end{aligned} \quad (10)$$

where $\varepsilon = 1/\epsilon$ and $\rho = \mu\gamma\varepsilon$. The last term of (10), named as reweighted zero attractor, provides a selective shrinkage to the taps. Due to this reweighted zero attractor, the inactive taps with zero magnitudes or magnitudes comparable to $1/\varepsilon$ undergo higher shrinkage compared to the active taps which enhances the performance both in terms of convergence speed and steady state EMSE.

III. CONVERGENCE ANALYSIS OF THE PROPOSED ZA-PNLMS ALGORITHM

A convergence analysis of the PNLMS algorithm is known to be a daunting task, due to the presence of $\mathbf{G}(n)$ both in the numerator and the denominator of the weight update term in (1), which again depends on $\mathbf{w}(n)$. The presence of the zero attractor term makes it further complicated for the proposed ZA-PNLMS algorithm, i.e., Eq. (9). To analyze the latter, we follow here an approach adopted recently in [26] in the context of PNLMS algorithm. This involves development of an equivalent transform domain model of the proposed algorithm first. A convergence analysis of the proposed algorithm is then carried out by applying to the equivalent model a scheme of angular discretization of continuous valued random vectors proposed first by Slock [22] and used later by several other researchers [24], [25].

A. A Transform Domain Model of the Proposed Algorithm

The proposed equivalent model uses a diagonal ‘transform’ matrix $\mathbf{G}^{\frac{1}{2}}(n)$ with $[\mathbf{G}^{\frac{1}{2}}(n)]_{i,i} = g_i^{\frac{1}{2}}(n)$, $i = 0, 1, \dots, L-1$, to transform the input vector $\mathbf{x}(n)$ and the filter coefficient vector $\mathbf{w}(n)$ to their ‘transformed’ versions, given respectively as $\mathbf{s}(n) = \mathbf{G}^{\frac{1}{2}}(n)\mathbf{x}(n)$ and $\mathbf{w}_N(n) = [\mathbf{G}^{\frac{1}{2}}(n)]^{-1}\mathbf{w}(n)$. It is easy to check that $\mathbf{w}_N^T(n)\mathbf{s}(n) = \mathbf{w}^T(n)\mathbf{x}(n) \equiv y(n)$ (say), i.e., the filter $\mathbf{w}_N(n)$ with input vector $\mathbf{s}(n)$ produces the same output $y(n)$ as produced by $\mathbf{w}(n)$ with input vector $\mathbf{x}(n)$. To compute $\mathbf{G}^{\frac{1}{2}}(n+1)$ and $\mathbf{w}_N(n+1)$, the filter $\mathbf{w}_N(n)$ is first updated to a weight vector $\mathbf{w}_N'(n+1)$ as

$$\begin{aligned} \mathbf{w}_N'(n+1) &= \mathbf{w}_N(n) + \frac{\mu e(n)\mathbf{s}(n)}{\mathbf{s}^T(n)\mathbf{s}(n) + \delta_P} \\ &\quad - \rho \mathbf{G}^{-\frac{1}{2}}(n) \text{sgn}(\mathbf{w}_N(n)). \end{aligned} \quad (11)$$

From (9), it is easy to check that $\mathbf{w}(n+1)$ is given by $\mathbf{w}(n+1) = \mathbf{G}^{\frac{1}{2}}(n)\mathbf{w}_N'(n+1)$. The matrix $\mathbf{G}(n+1)$ follows from $\mathbf{w}(n+1)$ following its definition and $\mathbf{w}_N(n+1)$ is then evaluated as $\mathbf{w}_N(n+1) = [\mathbf{G}^{\frac{1}{2}}(n+1)]^{-1}\mathbf{w}(n+1)$. From above, it follows that $\mathbf{w}_N(n+1) = \mathbf{G}^{-\frac{1}{2}}(n+1)\mathbf{w}(n+1) = \mathbf{G}^{-\frac{1}{2}}(n+1)\mathbf{G}^{\frac{1}{2}}(n)\mathbf{w}_N'(n+1)$, meaning $[\mathbf{w}_N(n+1)]_i = [\frac{g_i(n)}{g_i(n+1)}]^{\frac{1}{2}}[\mathbf{w}_N'(n+1)]_i$, $i = 0, 1, \dots, L-1$. Since $\sum_{i=0}^{L-1} g_i(n) = 1$ and $0 < g_i(n) < 1$, $i = 0, 1, \dots, L-1$, it is reasonable to expect that $g_i(n)$ does not change significantly from index n to index $(n+1)$ [especially near convergence and/or for large order filters] and thus, we can make the approximation $[g_i(n)]^{\frac{1}{2}}[\mathbf{w}_N(n+1)]_i \approx [g_i(n+1)]^{\frac{1}{2}}[\mathbf{w}_N(n+1)]_i$, which implies $\mathbf{w}_N(n+1) = \mathbf{w}_N(n+1)$.

B. Angular Discretization of a Continuous Valued Random Vector [22]

As per this, given a zero mean, $L \times 1$ random vector \mathbf{x} with correlation matrix $\mathbf{R} = E[\mathbf{x}\mathbf{x}^T]$, it is assumed that \mathbf{x} can assume only one of the $2L$ orthogonal directions, given by $\pm \mathbf{e}_i$, $i = 0, 1, \dots, L-1$, where \mathbf{e}_i is the i -th normalized eigenvector of \mathbf{R} corresponding to the eigenvalue λ_i . In particular, \mathbf{x} is modeled as $\mathbf{x} = s r \mathbf{v}$, where $\mathbf{v} \in \{\mathbf{e}_i | i = 0, 1, \dots, L-1\}$, with probability of $\mathbf{v} = \mathbf{e}_i$ given by p_i , $r = \|\mathbf{x}\|$, i.e., r has the same distribution as that of $\|\mathbf{x}\|$ and $s \in \{1, -1\}$, with probability of $s = \pm 1$ given by $P(s = \pm 1) = \frac{1}{2}$. Further, the three elements s , r and \mathbf{v} are assumed to be mutually independent. Note that as s is zero mean, $E[s r \mathbf{v}] = \mathbf{0}$ and thus $E[\mathbf{x}] = \mathbf{0}$ is satisfied trivially. To satisfy $E[\mathbf{x}\mathbf{x}^T] = \mathbf{R}$, the discrete probability p_i is taken as $p_i = \frac{\lambda_i}{\text{Tr}[\mathbf{R}]}$, which satisfies $p_i \geq 0$, $\sum_{i=0}^{L-1} p_i = 1$ and leads to $E[\mathbf{x}\mathbf{x}^T] = E(s^2 r^2 \mathbf{v}\mathbf{v}^T) = E(r^2) E(\mathbf{v}\mathbf{v}^T) = \text{Tr}[\mathbf{R}] \sum_{i=0}^{L-1} p_i \mathbf{e}_i \mathbf{e}_i^T = \sum_{i=0}^{L-1} \lambda_i \mathbf{e}_i \mathbf{e}_i^T = \mathbf{R}$. Also note that if θ_i be the angle between \mathbf{x} and \mathbf{e}_i , then $\cos(\theta_i) = \frac{\mathbf{x}^T \mathbf{e}_i}{\|\mathbf{x}\|}$ and $E[\cos^2(\theta_i)] \approx \frac{\lambda_i}{\text{Tr}[\mathbf{R}]}$, meaning p_i provides a measure of how far \mathbf{x} is (angularly) from \mathbf{e}_i on an average.

In our analysis of the proposed algorithm, we use the above model to represent the transformed input vector $\mathbf{s}(n)$ as

$$\mathbf{s}(n) = s_s(n) r_s(n) \mathbf{v}_s(n), \quad (12)$$

where, $s_s(n) \in \{+1, -1\}$ with $P(s_s(n) = \pm 1) = \frac{1}{2}$, $r_s(n) = \|\mathbf{s}(n)\|$ and $\mathbf{v}_s(n) \in \{\mathbf{e}_{s,i}(n) | i = 0, 1, \dots, L-1\}$ with $P(\mathbf{v}_s(n) = \mathbf{e}_{s,i}(n)) = \frac{\lambda_{s,i}(n)}{\text{Tr}[\mathbf{S}(n)]}$, where, $\mathbf{S}(n) = E[\mathbf{s}(n)\mathbf{s}^T(n)]$, $\lambda_{s,i}(n)$ is the i -th eigenvalue of $\mathbf{S}(n)$, and as before, the three elements $s_s(n)$, $r_s(n)$ and $\mathbf{v}_s(n)$ are mutually independent.

C. Convergence of the ZA-PNLMS Algorithm in mean

Now, defining the weight error vector at the n -th index as $\tilde{\mathbf{w}}(n) = \mathbf{w}_{opt} - \mathbf{w}(n)$, the transform domain weight error vector $\tilde{\mathbf{w}}_N(n) = \mathbf{G}^{-\frac{1}{2}}(n) \tilde{\mathbf{w}}(n) \equiv \mathbf{G}^{-\frac{1}{2}}(n) \mathbf{w}_{opt} - \mathbf{w}_N(n)$ and expressing $e(n) = \mathbf{s}^T(n) \tilde{\mathbf{w}}_N(n) + v(n)$, the recursive form of the weight error vectors can then be obtained as

$$\begin{aligned} \tilde{\mathbf{w}}_N(n+1) &\approx \tilde{\mathbf{w}}_N(n) - \frac{\mu \mathbf{s}(n) \mathbf{s}^T(n) \tilde{\mathbf{w}}_N(n)}{\mathbf{s}^T(n) \mathbf{s}(n) + \delta_P} \\ &\quad - \frac{\mu \mathbf{s}(n) v(n)}{\mathbf{s}^T(n) \mathbf{s}(n) + \delta_P} + \rho \mathbf{G}^{-\frac{1}{2}}(n) \text{sgn}(\mathbf{w}_N(n)). \end{aligned} \quad (13)$$

For our analysis here, we approximate δ_P by zero in (13) as δ_P is a very very small constant. The first order convergence of the ZA-PNLMS is then provided in the following theorem.

Theorem 1. *With a zero-mean input $x(n)$ of covariance matrix \mathbf{R} , the ZA-PNLMS algorithm produces stable $\tilde{\mathbf{w}}_N(n)$ and also $\tilde{\mathbf{w}}(n)$ if the step-size μ satisfies $0 < \mu < 2$ and under this condition, $\tilde{\mathbf{w}}_N(n)$ and $\tilde{\mathbf{w}}(n)$ converge respectively as per the following:*

$$\begin{aligned} \tilde{\mathbf{w}}_N(\infty) &= \lim_{n \rightarrow \infty} \tilde{\mathbf{w}}_N(n) = E(\mathbf{G}^{-\frac{1}{2}}(n))|_{\infty} \mathbf{w}_{opt} \\ &\quad - \frac{\rho}{\mu} \text{Tr}(\mathbf{S}(\infty)) \mathbf{S}^{-1}(\infty) \lim_{n \rightarrow \infty} E(\mathbf{G}^{-\frac{1}{2}}(n) \text{sgn}(\mathbf{w}_N(n))) \end{aligned} \quad (14)$$

and

$$\begin{aligned} \tilde{\mathbf{w}}(\infty) &= \lim_{n \rightarrow \infty} \tilde{\mathbf{w}}(n) = \mathbf{w}_{opt} - \frac{\rho}{\mu} \text{Tr}(\mathbf{S}(\infty)) \times \\ &\quad E(\mathbf{G}^{\frac{1}{2}}(n))|_{\infty} \mathbf{S}^{-1}(\infty) \lim_{n \rightarrow \infty} E(\mathbf{G}^{-\frac{1}{2}}(n) \text{sgn}(\mathbf{w}(n))), \end{aligned} \quad (15)$$

where $\mathbf{S}(n) = E(\mathbf{s}(n)\mathbf{s}^T(n)) = E(\mathbf{G}^{\frac{1}{2}}(n) \mathbf{R} \mathbf{G}^{\frac{1}{2}}(n))$.

Proof: For analysis, we now substitute $\delta = 0$ in (13) as δ is a very very small constant. Taking expectation of both sides of (13) and invoking the well known “independence assumption” that allows taking $\mathbf{w}_N(n)$ to be statistically independent of $\mathbf{s}(n)$, we then obtain,

$$\begin{aligned} E(\tilde{\mathbf{w}}_N(n+1)) &= E(\tilde{\mathbf{w}}_N(n)) - \mu E\left(\frac{\mathbf{s}(n) \mathbf{s}^T(n)}{\mathbf{s}^T(n) \mathbf{s}(n)}\right) E(\tilde{\mathbf{w}}_N(n)) + \rho E(\mathbf{G}^{-\frac{1}{2}}(n) \text{sgn}(\mathbf{w}_N(n))) \\ &\Rightarrow E(\tilde{\mathbf{w}}_N(n+1)) = (\mathbf{I} - \mu \mathbf{B}(n)) E(\tilde{\mathbf{w}}_N(n)) + \rho E(\mathbf{G}^{-\frac{1}{2}}(n) \text{sgn}(\mathbf{w}_N(n))) \end{aligned} \quad (16)$$

where

$$\mathbf{B}(n) = E\left(\frac{\mathbf{s}(n) \mathbf{s}^T(n)}{\mathbf{s}^T(n) \mathbf{s}(n)}\right). \quad (17)$$

Note that $\mathbf{B}(n)$ is symmetric and therefore, one can have its eigendecomposition $\mathbf{B}(n) = \mathbf{E}(n)\mathbf{D}(n)\mathbf{E}^T(n)$ where $\mathbf{E}(n) = [\mathbf{e}_0(n) \mathbf{e}_1(n) \cdots \mathbf{e}_{L-1}(n)]$, $\mathbf{D}(n) = \text{diag}[\lambda_0(n), \lambda_1(n), \cdots, \lambda_{L-1}(n)]$, with $\mathbf{e}_i(n)$ and $\lambda_i(n)$, $i = 0, 1, \cdots, L-1$ denoting the i -th eigenvector and eigenvalue of $\mathbf{B}(n)$ respectively. The eigenvalues are real and the eigenvectors $\mathbf{e}_i(n)$ are mutually orthonormal, meaning $\mathbf{E}(n)$ is unitary, i.e., $\mathbf{E}^T(n)\mathbf{E}(n) = \mathbf{E}(n)\mathbf{E}^T(n) = \mathbf{I}$. From $\mathbf{B}(n)\mathbf{e}_i(n) = \lambda_i(n)\mathbf{e}_i(n)$ and the fact that $\|\mathbf{e}_i(n)\|^2 = 1$, it is easy to observe that

$$\lambda_i(n) = \mathbf{e}_i^T(n)\mathbf{B}(n)\mathbf{e}_i(n) = E \left[\frac{[\mathbf{s}^T(n)\mathbf{e}_i(n)]^2}{\|\mathbf{s}(n)\|^2} \right].$$

Two observations can be made now:

- 1) $\lambda_i(n) > 0$ [theoretically, one can have $\lambda_i(n) = 0$ also, provided $\mathbf{s}^T(n)\mathbf{e}_i(n) = 0$, i.e., $\mathbf{s}(n)$ is orthogonal to $\mathbf{e}_i(n)$ in each trial, which is ruled out here].
- 2) From Cauchy-Schwarz inequality, $[\mathbf{s}^T(n)\mathbf{e}_i(n)]^2 \leq \|\mathbf{s}(n)\|^2 \|\mathbf{e}_i(n)\|^2 = \|\mathbf{s}(n)\|^2$, meaning $\lambda_i(n) \leq 1$.

Pre-multiplying both sides of (16) by $\mathbf{E}^T(n)$, defining $\mathbf{u}(n) = \mathbf{E}^T(n)E(\tilde{\mathbf{w}}_N(n))$, $\mathbf{v}(n+1) = \mathbf{E}^T(n)E(\tilde{\mathbf{w}}_N(n+1))$, $\mathbf{z}(n) = \mathbf{E}^T(n)E(\mathbf{G}^{-\frac{1}{2}}(n)\text{sgn}(\mathbf{w}_N(n)))$, substituting $\mathbf{B}(n)$ by $\mathbf{E}(n)\mathbf{D}(n)\mathbf{E}^T(n)$ and using the unitariness of $\mathbf{E}(n)$, we have,

$$\mathbf{v}(n+1) = (\mathbf{I} - \mu\mathbf{D}(n))\mathbf{u}(n) + \rho\mathbf{z}(n). \quad (18)$$

Taking norm on both sides of (18) and invoking triangle inequality property of norm, i.e., $\|\mathbf{a} + \mathbf{b}\| \leq \|\mathbf{a}\| + \|\mathbf{b}\|$, we then obtain,

$$\|\mathbf{v}(n+1)\| \leq \|(\mathbf{I} - \mu\mathbf{D}(n))\mathbf{u}(n)\| + \rho\|\mathbf{z}(n)\|. \quad (19)$$

Since $\mathbf{E}(n)$ is unitary, we have $\|\mathbf{v}(n+1)\| = \|E(\tilde{\mathbf{w}}_N(n+1))\|$, $\|\mathbf{u}(n)\| = \|E(\tilde{\mathbf{w}}_N(n))\|$ and $\|\mathbf{z}(n+1)\| = \left\| E \left[\mathbf{G}^{-\frac{1}{2}}(n)\text{sgn}(\mathbf{w}_N(n)) \right] \right\|$. Using the fact that $\{E[g_{ii}^{-\frac{1}{2}}(n)\text{sgn}(\mathbf{w}_N(n))]\}^2 \leq E(g_{ii}^{-1}(n))$ (i.e., using Cauchy Schwarz inequality and the fact that $\text{sgn}^2(\cdot) = 1$), we can write $\|\mathbf{z}(n+1)\| \leq \sqrt{\sum_{i=0}^{L-1} E(g_{ii}^{-1}(n))} = c(n)$ (say). Clearly $c(n)$ is finite, as $\mathbf{G}(n)$ is a diagonal matrix with only positive elements. From (19), one can then write,

$$\|E(\tilde{\mathbf{w}}_N(n+1))\| \leq \sqrt{\sum_{i=0}^{L-1} (1 - \mu\lambda_i(n))^2 |u_i(n)|^2} + \rho c(n). \quad (20)$$

We now select μ so that $|1 - \mu\lambda_i(n)| < 1$, or, equivalently, $-1 < 1 - \mu\lambda_i(n) < 1$, which leads to the following:

- 1) $\mu\lambda_i(n) > 0$, meaning $\mu > 0$ as $\lambda_i(n) > 0$ (as explained above).
- 2) $\mu < \frac{2}{\lambda_i(n)}$ (since $\lambda_i(n) \leq 1$ or, equivalently $\frac{1}{\lambda_i(n)} \geq 1$, it will be sufficient to take $\mu < 2$ for satisfying this inequality).

Therefore, for $0 < \mu < 2$, we have $|1 - \mu\lambda_i(n)| < 1$, where $0 < \lambda_i(n) \leq 1$. Let $\|E(\tilde{\mathbf{w}}_N(n))\| = \theta(n)$ and $k(n) = \max \{|(1 - \mu\lambda_i(n))|, i = 0, 1, \cdots, L-1\}$, meaning $0 \leq k(n) < 1$. From (20), one can then write,

$$\theta(n+1) \leq k(n) \sqrt{\sum_{i=0}^{L-1} |u_i(n)|^2} + \rho c(n) = k(n)\theta(n) + \rho c(n). \quad (21)$$

Proceeding recursively backwards till $n = 0$,

$$\theta(n+1) \leq \prod_{i=0}^n k(n-i) \theta(0) + \rho \left(c(n) + \sum_{l=0}^{n-1} \left(\prod_{i=0}^l k(n-i) \right) c(n-l-1) \right). \quad (22)$$

Clearly, for $0 \leq k(n) < 1$, the first term of RHS of (22) vanishes as n approaches infinity. For the second term, $c(n)$ is a bounded sequence, which, in steady state, can be taken to be time invariant, say c , as the variation of $g_{ii}(n)$ vs. n , $i = 0, 1, \cdots, L-1$ in the steady state are negligible. Also note that $\prod_{i=0}^l k(n-i)$ is a decaying function of l since $0 \leq k(m) < 1$ at any index m . From these, one can write $\lim_{n \rightarrow \infty} \|E(\tilde{\mathbf{w}}_N(n))\| = \lim_{n \rightarrow \infty} \theta(n) \leq \rho K$ where K is a positive constant. Recalling that $\tilde{\mathbf{w}}(n) = \mathbf{G}^{\frac{1}{2}}(n)\tilde{\mathbf{w}}_N(n)$, we can then write $\lim_{n \rightarrow \infty} \|E(\tilde{\mathbf{w}}(n))\| \approx \lim_{n \rightarrow \infty} \left\| E(\mathbf{G}^{\frac{1}{2}}(n))E(\tilde{\mathbf{w}}_N(n)) \right\| < \lim_{n \rightarrow \infty} \|E(\tilde{\mathbf{w}}_N(n))\| \leq \rho K$. Since ρ is very small, this implies that $E(\tilde{\mathbf{w}}(n))$ will remain in close vicinity of \mathbf{w}_{opt} in the steady state under the condition: $0 < \mu < 2$. In other words, $E(\tilde{\mathbf{w}}(n))$ will provide a biased estimate of \mathbf{w}_{opt} , though the bias, being proportional to ρ , will be negligibly small.

Under the condition $0 < \mu < 2$, letting n approach infinity on both the LHS and the RHS of (16) and noting that as $n \rightarrow \infty$, $E(\tilde{\mathbf{w}}_N(n+1)) \approx E(\tilde{\mathbf{w}}_N(n))$, one can obtain from (16),

$$\lim_{n \rightarrow \infty} E(\tilde{\mathbf{w}}_N(n)) = \frac{\rho}{\mu} \mathbf{B}^{-1}(\infty) \lim_{n \rightarrow \infty} E(\mathbf{G}^{-\frac{1}{2}}(n) \times \text{sgn}(\mathbf{G}^{\frac{1}{2}}(n)\mathbf{w}_N(n))). \quad (23)$$

Further, $\mathbf{B}(n)$ can be simplified in terms of $\mathbf{S}(n)$ by invoking the angular discretization model of a random vector as discussed in the section III.B. We replace $\mathbf{s}(n)$ by $s_s(n)r_s(n)\mathbf{v}_s(n)$ as given by (12). One can then write,

$$\begin{aligned}\mathbf{B}(n) &= E\left(\frac{\mathbf{s}(n)\mathbf{s}^T(n)}{\mathbf{s}^T(n)\mathbf{s}(n)}\right) = E\left(\frac{r_s^2(n)\mathbf{v}_s(n)\mathbf{v}_s^T(n)}{r_s^2(n)\mathbf{v}_s^T(n)\mathbf{v}_s(n)}\right) \quad (\text{since } s_s^2(n) = 1) \\ &= \sum_{i=0}^{L-1} \frac{\lambda_{s,i}(n)}{Tr(\mathbf{S}(n))} \mathbf{e}_{s,i}(n) \mathbf{e}_{s,i}^T(n) = \frac{\mathbf{S}(n)}{Tr(\mathbf{S}(n))},\end{aligned}\quad (24)$$

since $\mathbf{S}(n) = \sum_{i=0}^{L-1} \lambda_{s,i}(n) \mathbf{e}_{s,i}(n) \mathbf{e}_{s,i}^T(n)$.

Letting n approach infinity in (24) and substituting this in (23),

$$\lim_{n \rightarrow \infty} E(\tilde{\mathbf{w}}_N(n)) = \frac{\rho}{\mu} Tr(\mathbf{S}(\infty)) \mathbf{S}^{-1}(\infty) \lim_{n \rightarrow \infty} E(\mathbf{G}^{-\frac{1}{2}}(n) \times sgn(\mathbf{G}^{\frac{1}{2}}(n) \mathbf{w}_N(n))). \quad (25)$$

Recalling $\tilde{\mathbf{w}}(n) = \mathbf{G}^{\frac{1}{2}}(n) \tilde{\mathbf{w}}_N(n)$, from (25) we have,

$$\begin{aligned}\lim_{n \rightarrow \infty} E(\tilde{\mathbf{w}}(n)) &\approx \lim_{n \rightarrow \infty} E(\mathbf{G}^{\frac{1}{2}}(n)) E(\tilde{\mathbf{w}}_N(n)) \\ &= \frac{\rho}{\mu} Tr(\mathbf{S}(\infty)) E(\mathbf{G}^{\frac{1}{2}}(n)) \Big|_{\infty} \mathbf{S}^{-1}(\infty) \lim_{n \rightarrow \infty} E(\mathbf{G}^{-\frac{1}{2}}(n) sgn(\mathbf{w}(n))).\end{aligned}\quad (26)$$

Further, we have $E(\mathbf{w}_N(n)) = E(\mathbf{G}^{-\frac{1}{2}}(n) \mathbf{w}_{opt}) - E(\tilde{\mathbf{w}}_N(n))$ and $E(\mathbf{w}(n)) = \mathbf{w}_{opt} - E(\tilde{\mathbf{w}}(n))$, and thus, with (25) and (26), this completes the proof. ■

Corollary 1. For white input, $\bar{w}_i(\infty) (= \lim_{n \rightarrow \infty} E(w_i(n)))$ for the i -th active tap (i.e., for which $w_{opt,i} \neq 0$) is approximately given by

$$\bar{w}_i(\infty) = w_{opt,i} - \frac{\rho}{\mu} \bar{g}_i^{-1}(\infty) sgn(w_{opt,i}) \quad (27)$$

where $\bar{g}_i(\infty) = \lim_{n \rightarrow \infty} \bar{g}_i(n)$ and $\bar{g}_i(n) = [E(\mathbf{G}(n))]_{i,i}$.

Proof: For white input with variance σ_x^2 , we have $\mathbf{R} = \sigma_x^2 \mathbf{I}$, $\mathbf{S}(n) = \sigma_x^2 E(\mathbf{G}(n))$, $Tr(\mathbf{S}(n)) = \sigma_x^2$ and $\mathbf{S}^{-1}(n) = \frac{1}{\sigma_x^2} E(\mathbf{G}(n))^{-1}$ and then, we can have a simplified expression of $\bar{\mathbf{w}}(\infty)$ as

$$\bar{\mathbf{w}}(\infty) \approx \mathbf{w}_{opt} - \frac{\rho}{\mu} \lim_{n \rightarrow \infty} E(\mathbf{G}(n))^{-1} E(sgn(\mathbf{w}(n))) \quad (28)$$

where we have assumed that in the steady state as $n \rightarrow \infty$, $\mathbf{G}^{-\frac{1}{2}}(n)$ and $sgn(\mathbf{w}(n))$ become statistically independent and $E(\mathbf{G}(n)^{-\frac{1}{2}}) \approx E(\mathbf{G}(n))^{-\frac{1}{2}}$, which is reasonable as in the steady state, variance of each individual $g_i(n)$, $i = 0, 1, \dots, L-1$ is quite small (i.e., it behaves almost like a constant). Now, for an active tap with significantly large magnitude $w_{opt,i}$, it is reasonable to approximate $sgn(w_i(n)) \approx sgn(w_{opt,i})$ under the assumption that in the steady state, the variance of $w_i(n)$, i.e., $E((w_i(n) - w_{opt,i})^2)$ is small enough compared to the magnitude of $w_{opt,i}$. Then, with $E(sgn(w_i(n))) \approx E(sgn(w_{opt,i})) = sgn(w_{opt,i})$ for an active tap in the steady state, the result follows trivially from (28). ■

Corollary 1 shows that

$$\bar{w}_i(\infty) = \begin{cases} w_{opt,i} - \frac{\rho}{\mu} \bar{g}_i^{-1}(\infty), & \text{if } sgn(w_{opt,i}) > 0 \\ w_{opt,i} + \frac{\rho}{\mu} \bar{g}_i^{-1}(\infty), & \text{if } sgn(w_{opt,i}) < 0, \end{cases}$$

which implies that $\bar{w}_i(\infty)$ is always closer to the origin vis-a-vis $w_{opt,i}$. Further, the bias (i.e., usually defined as $w_{opt,i} - \bar{w}_i(\infty)$) is also proportional to $\bar{g}_i^{-1}(\infty)$, meaning active taps with comparatively smaller values will have larger bias and vice versa.

In the case of inactive taps, we have $w_{opt,i} = 0$. From (14) and for $\rho = 0$ (i.e., no zero attraction), this implies $\bar{w}_i(\infty) = 0$, i.e., the tap estimates fluctuate around zero value. For $\rho > 0$, the zero attractors come into play in the update equation (7) and act as an additional force that tries to pull the coefficients to zero from either side. The effect of zero attractor is thus to confine the fluctuations in a small band around zero. On an average, one can then take $E(sgn(w_i(n))) \Big|_{\infty} \approx 0$, meaning, from (16), the inactive tap estimates will largely be free of any bias.

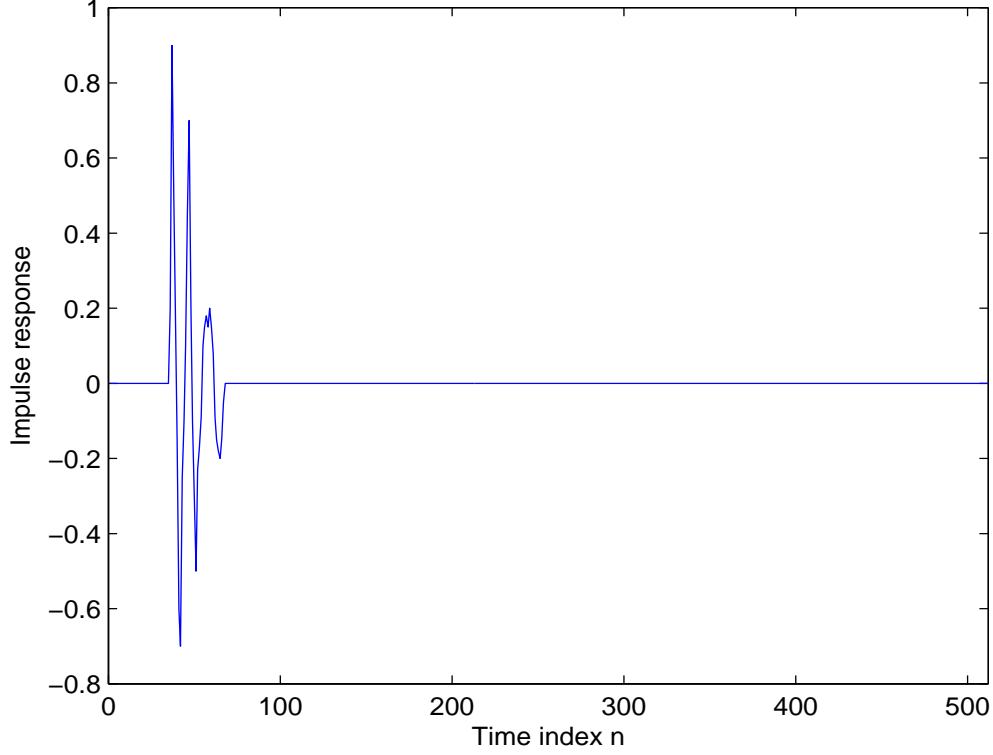


Fig. 1. Impulse response of the sparse system

IV. NUMERICAL SIMULATIONS

In this section, we investigate evolution of $E(\mathbf{w}(n))$ of the proposed ZA-PNLMS algorithm with time via simulation studies in the context of sparse system identification. For this, we considered a sparse system with impulse responses of length $L=512$ as shown in Fig. 1. The system has 37 active taps and is driven by a zero mean, white input $x(n)$ of variance $\sigma_x^2 = 1$, with the output observation noise $v(n)$ being taken to be zero mean, white Gaussian with $\sigma_v^2 = 10^{-3}$. The proposed ZA-PNLMS algorithm is used to identify the system, for which the step size μ , the zero attracting coefficient ρ and the regularization parameter (to avoid division by zero) are taken to be 0.7, 0.0001 and 0.01 respectively, while ρ_g and δ are chosen as 0.01 and 0.001 respectively. The simulations are carried out for a total of 25,000 iterations and for each tap weight $w_i(n)$, the learning curve $E[w_i(n)]$ vs n is evaluated by averaging $w_i(n)$ over 30 experiments. For demonstration here, we consider four representative learning curves, for $i=37, 55, 67, 1$. (the corresponding $w_{opt,i}$ given by 0.9, 0.1, -0.05 and 0 respectively). These are shown in Figs. 2-5 respectively where it is seen that for both the inactive tap (i.e., $w_{opt,1}$) and the active tap with relatively large magnitude (i.e., $w_{opt,37}(n)$), $E[w_i(n)]$ converges to its optimum values of 0 and 0.9 respectively. On the other hand, for $w_{opt,67}(n)$ and $w_{opt,55}(n)$, i.e., for active taps with relatively less magnitudes, $E[w_i(n)]$ converges with reasonably large bias. This validates our conjectures made in section III (Corollary 1 and the subsequent analysis). To validate the same further, the bias is calculated from the learning curves (in the steady state) for all the taps and then plotted in Fig. 6 against the magnitude of the optimum tap weights. Clearly, the bias becomes negligible as the magnitude of the active tap increases.

REFERENCES

- [1] J. Radecki, Z. Zilic, and K. Radecka, "Echo cancellation in IP networks", *Proceedings of the 45th Midwest Symposium on Circuits and Systems*, vol. 2, pp. 219-222, Tulsa, Okla, USA, August 2002.
- [2] V. V. Krishna, J. Rayala and B. Slade, "Algorithmic and Implementation Aspects of Echo Cancellation in Packet Voice Networks", *Proc. Thirty-Sixth Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 1252-1257, 2002.
- [3] E. Hansler and G. Schmidt, Eds., *Topics in Acoustic Echo and Noise Control*, Berlin, Springer-Verlag, 2006.
- [4] W. Schreiber, "Advanced Television Systems for Terrestrial Broadcasting", *Proc. IEEE*, vol. 83, no. 6, pp. 958-981, 1995.
- [5] W. Bajwa, J. Haupt, G. Raz and R. Nowak, "Compressed Channel Sensing", *Prof. IEEE CISS*, 2008, pp. 5-10.
- [6] M. Kocic, D. Brady and M. Stojanovic, "Sparse Equalization for Real-Time Digital Underwater Acoustic Communications", *Proc. IEEE OCEANS*, 1995, pp. 1417-1422.
- [7] R. L. Das and M. Chakraborty, "Sparse Adaptive Filters - an Overview and Some New Results", *Proc. ISCAS-2012*, pp. 1267-1270, COEX, Seoul, Korea, May 20-23, 2012.

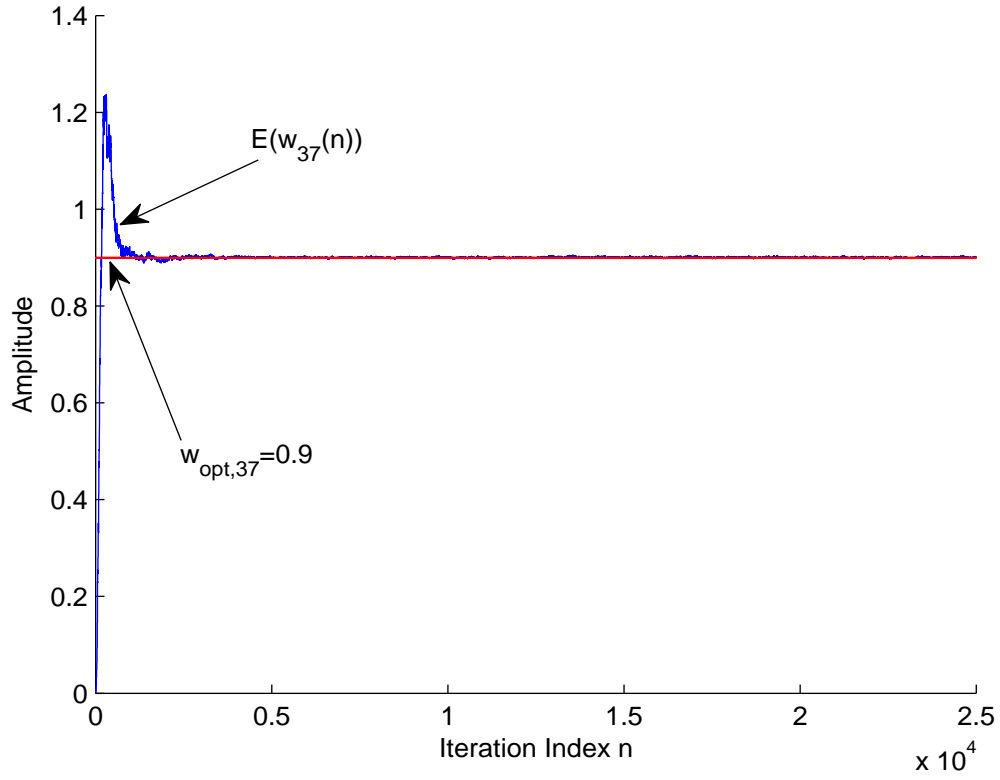


Fig. 2. Evolution of $E(w_{37}(n))$ of the ZA-PNLMS algorithm with its optimum level $w_{opt,37} = 0.9$.

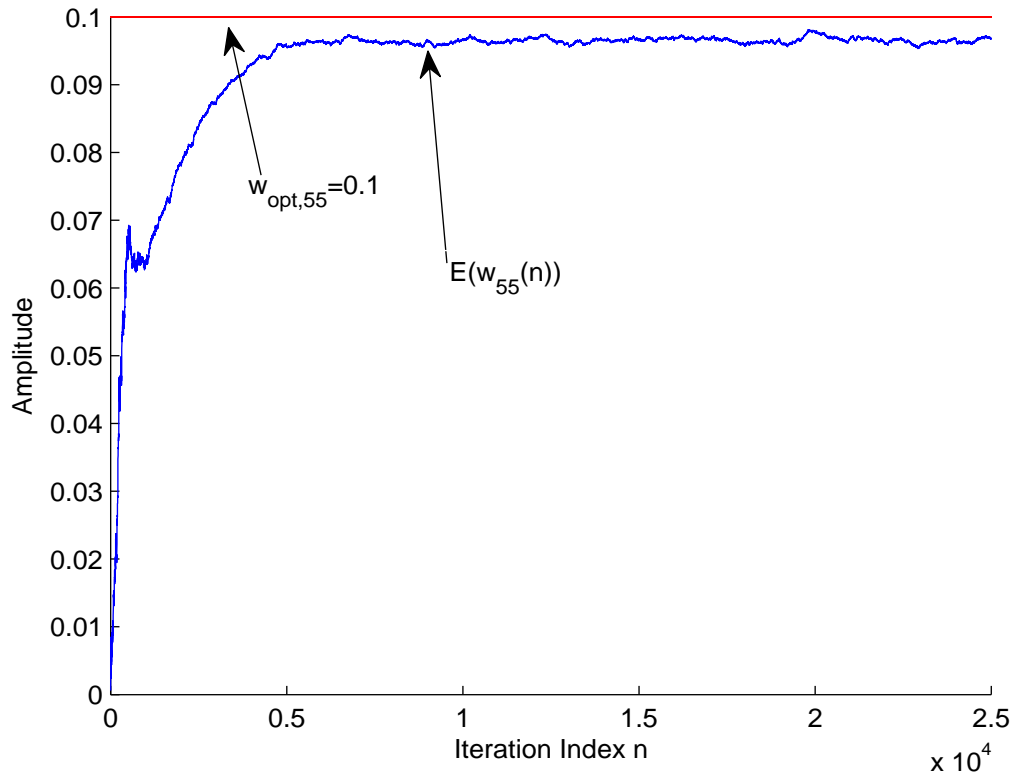


Fig. 3. Evolution of $E(w_{55}(n))$ of the ZA-PNLMS algorithm with its optimum level $w_{opt,55} = 0.1$.

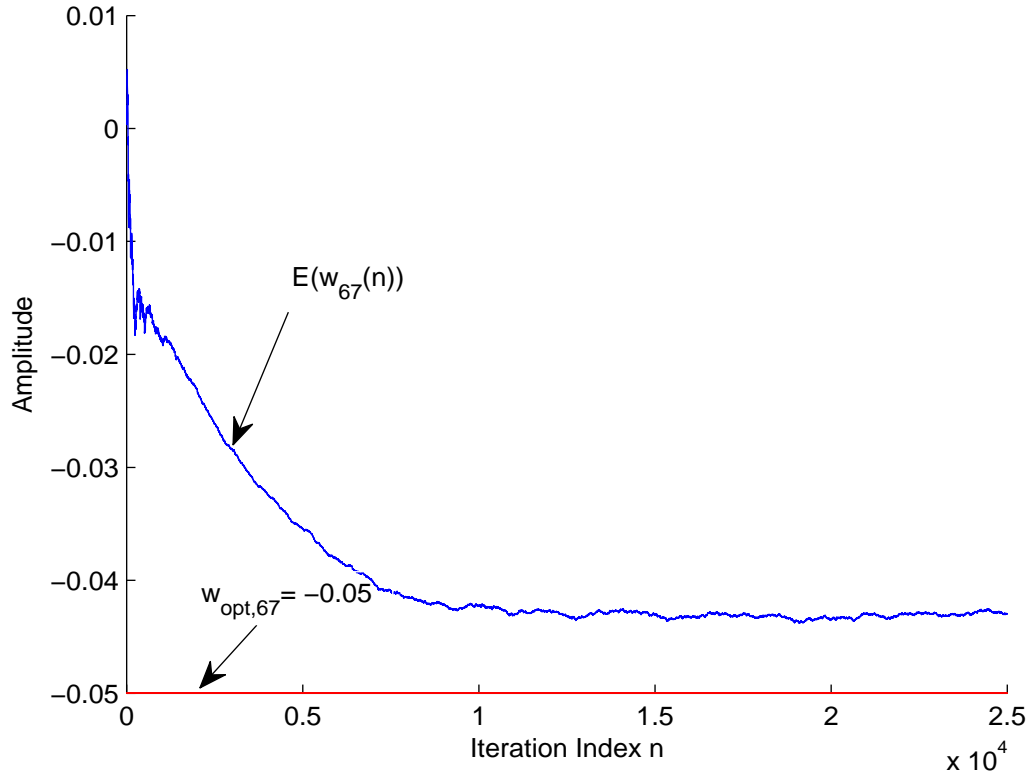


Fig. 4. Evolution of $E(w_{67}(n))$ of the ZA-PNLMS algorithm with its optimum level $w_{opt,67} = -0.5$.

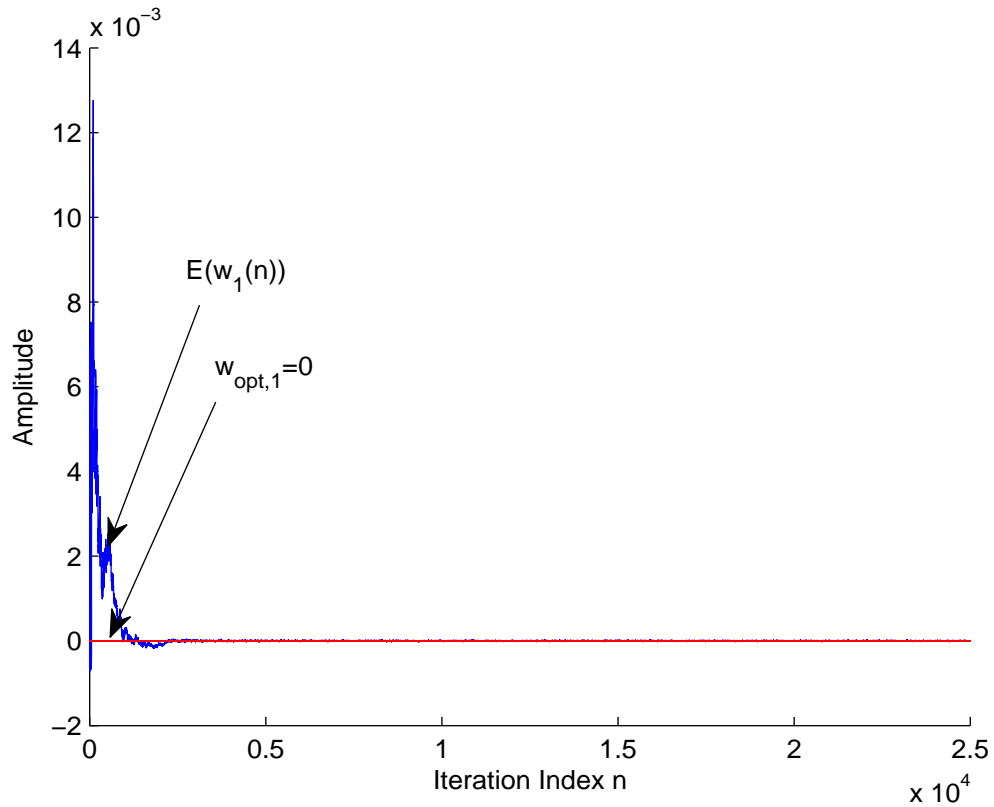


Fig. 5. Evolution of $E(w_1(n))$ of the ZA-PNLMS algorithm with its optimum level $w_{opt,1} = 0$.

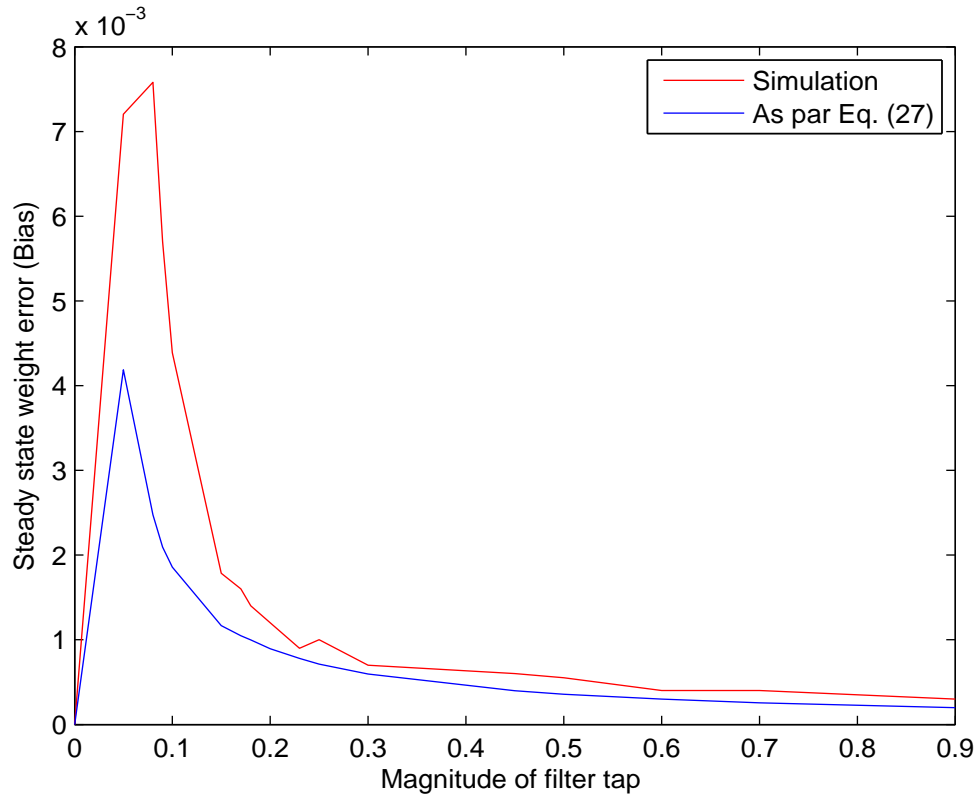


Fig. 6. Evolution of $E(w_1(n))$ of the ZA-PNLMS algorithm with its optimum level $w_{opt,1} = 0$.

- [8] D.L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers", *IEEE Trans. Speech Audio Process.*, vol. 8, no. 5, pp. 508-518, September 2000.
- [9] S. Haykin, *Adaptive Filter Theory*, Englewood Cliffs, NJ: Prentice-Hall, 1986.
- [10] S.L. Gay, "An efficient, fast converging adaptive filter for network echo cancellation", *Proc. Asilomar Conf. Signals, Systems, Comput.*, pp. 394-398, Nov., 1998.
- [11] J. Benesty and S.L. Gay, "An improved PNLMS algorithm", *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing*, pp. 1881-1884, 2002, Orlando, Florida, USA.
- [12] M. A. Mehran Nekui, "A Fast Converging Algorithm for Network Echo Cancellation", *IEEE Signal Processing Letters*, vol. 11, no. 4, pp. 427-430, April, 2004.
- [13] H. Deng and M. Doroslovacki, "Improving convergence of the PNLMS algorithm for sparse impulse response identification", *IEEE Signal Processing Letters*, vol. 12, no. 3, pp. 181-184, 2005.
- [14] L. Liu, M. Fukumoto, S. Saiki and S. Zhang, "A Variable Step-Size Proportionate Affine Projection Algorithm for Identification of Sparse Impulse Response", *EURASIP Journal on Advances in Signal Processing*, pp. 1-10, 2009.
- [15] L. Liu, M. Fukumoto, S. Saiki and S. Zhang, "A Variable Step-Size Proportionate NLMS Algorithm for Identification of Sparse Impulse Response", *IEICE Trans. Fundamentals*, vol. E93A, no. 1, Jan. 2010.
- [16] C. Paleologu, J. Benesty, F. Albu and S. Ciochină, "An Efficient Variable Step-Size Proportionate Affine Projection Algorithm", *Proc. IEEE ICASSP*, pp. 77-80, 2011.
- [17] Y. Chen, Y. Gu, and A.O. Hero, "Sparse LMS for system identification", *Proc. IEEE ICASSP-2009*, pp. 3125-3128, Apr. 2009, Taipei, Taiwan.
- [18] Y. Chen, Y. Gu, and A.O. Hero, "Regularized Least-Mean-Square Algorithms", *Arxiv preprint arXiv:1012.5066v2[stat.ME]*, Dec. 2010.
- [19] R. L. Das and M. Chakraborty, "A Zero Attracting Proportionate Normalized Least Mean Square Algorithm", *Proc. APSIPA-ASC-2012*, pp. 1-4, Hollywood, California, 3-6 Dec. 2012.
- [20] R. L. Das and M. Chakraborty, "A Variable Step-Size Zero Attracting Proportionate Normalized Least Mean Square Algorithm", *Proc. ISCAS-2014*, COEX, Melbourne, Australia, June 1-5, 2014.
- [21] M. Yukawa and I. Yamada, "A unified view of adaptive variable - metric projection algorithms", *EURASIP Journal on Advances in Signal Processing*, vol. 2009, pp. 1-13, 2009.
- [22] D. T. M. Slock, "On the convergence behavior of the lms and the normalized lms algorithms", *IEEE Trans. on Signal Processing*, vol. 41, no. 9, pp. 2811-2825, Sep. 1993.
- [23] Z. Yang, Y. R. Zheng and S. L. Grant, "Proportionate affine projection sign algorithms for network echo cancellation", *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 19, no. 8, pp. 2273-2284, Nov. 2011.
- [24] S. G. Sankaran and A. A. (Louis) Beex, "Convergence behavior of affine projection algorithms", *IEEE Trans. on Signal Processing*, vol. 48, no. 4, pp. 1086-1096, Apr. 2000.
- [25] T. K. Paul and T. Ogunfunmi, "On the convergence behavior of the affine projection algorithm for adaptive filters", *IEEE Trans. on Circuits and System-1:Regular Papers*, vol. 58, no. 8, pp. 1813-1826, Aug. 2011.
- [26] R. L. Das and M. Chakraborty, "On Convergence of Proportionate-type Normalized Least Mean Square Algorithms", *IEEE Trans. Circuits and Systems II: Express Briefs*, vol. 62, no. 5, pp. 491-495, May, 2015.
- [27] R. Price, "A useful theorem for nonlinear devices using gaussian inputs", *IRE Trans. on Information Theory*, vol. 4, no. 2, pp. 6972, Jun. 1958.
- [28] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. New York, NY, USA: McGraw Hill, 2002.

- [29] R. H. Kwong and E. W. Johnston, "A variable step size LMS algorithm", *IEEE Trans. Signal Processing*, vol. 40, no. 7, pp. 1633-1642, July 1992.